North American Journal of Economics and Finance 39 (2017) 38–55

Contents lists available at ScienceDirect

North American Journal of Economics and Finance

journal homepage: www.elsevier.com/locate/ecofin

Pair trading based on quantile forecasting of smooth transition GARCH models

Cathy W.S. Chen [a,*], Zona Wang [a], Songsak Sriboonchitta [b], Sangyeol Lee [c]

[a] Department of Statistics, Feng Chia University, Taiwan
[b] Faculty of Economics, Chiang Mai University, Thailand
[c] Department of Statistics, Seoul National University, Republic of Korea

ARTICLE INFO

Article history:
Received 30 May 2016
Received in revised form 18 October 2016
Accepted 19 October 2016
Available online 10 November 2016

JEL Classification:
C11
C22
C51
C53
C58

Keywords:
Pair trading
Bayesian inference
Smooth transition GARCH model
Second-order logistic transition function
Markov chain Monte Carlo methods
Out-of-sample forecasts
Quantile forecasting

ABSTRACT

Pair trading is a statistical arbitrage strategy used on similar assets with dissimilar valuations. We utilize smooth transition heteroskedastic models with a second-order logistic function to generate trading entry and exit signals and suggest two pair trading strategies: the first uses the upper and lower threshold values in the proposed model as trading entry and exit signals, while the second strategy instead takes one-step-ahead quantile forecasts obtained from the same model. We employ Bayesian Markov chain Monte Carlo sampling methods for updating the estimates and quantile forecasts. As an illustration, we conduct a simulation study and empirical analysis of the daily stock returns of 36 stocks from U.S. stock markets. We use the minimum square distance method to select ten stock pairs, choose additional five pairs consisting of two companies in the same industrial sector, and then finally consider pair trading profits for two out-of-sample periods in 2014 within a six-month time frame as well as for the entire year. The proposed strategies yield average annualized returns of at least 35.5% without a transaction cost and at least 18.4% with a transaction cost.

1. Introduction

Pair trading is a quantitative method of speculation, pioneered by Nunzio Tartaglia's quantitative analysis group at Morgan Stanley in the 1980s. They used sophisticated statistical techniques to design a trading program, and one of the strategies was to identify pairs of securities whose prices tend to move together. Pair trading is a mean-reverting strategy that assumes the spread calculated from the two stock returns, i.e. the difference between the two prices, will revert to its historical trend, thus having the capacity to achieve profits through simple and relatively low-risk positions. It is also a market-neutral trading strategy that matches a long position with a short position through a pair of highly correlated instruments, such as two stocks, exchange-traded funds (ETFs), currencies, commodities, options, etc. A profit or loss on a pair trade depends on whether the spread between the paired positions widens or narrows. When the spread between the stocks widens, traders

* Corresponding author.
E-mail address: chenws@mail.fcu.edu.tw (C.W.S. Chen).

http://dx.doi.org/10.1016/j.najef.2016.10.015
1062-9408/

sell the higher priced stock and buy the lower priced stock. Trades such as these will result in profit if the two stock prices converge.

The basic idea of pair trading is to take advantage of market inefficiencies. There are many pair trading selection methods in the literature, which can be classified in four ways: distance method, cointegration method, time-series method, and stochastic spread method. Krauss (2016) comprehensively reviews the literature on pair trading frameworks, categorizing them into five groups, including an extra one that is entitled "other approaches". Regarding the distance method, Gatev, Goetzmann, and Rouwenhorst (2006) propose a pair trading strategy using relative stock price movements as triggers to open or close a pair position. Their commonly-used procedure measures the distance, or the sum of squared differences, between the two normalized price series, wherein the trading trigger is the two historical standard deviations estimated during the formation period.

The cointegration method, detailed in Vidyamurthy (2004), attempts to relate the cointegration model to the Arbitrage Pricing Theory (APT) (Ross, 1976). Perlin (2009) uses the cointegration method to find stock pairs in the Brazil stock market. Pair selection is executed by maximizing the Pearson correlation between standardized price time series, which is equivalent to minimizing the sum of the squared distance. Girma and Paulson (1999) propose to employ the price difference between petroleum futures and futures on its refined end products from 1983 to 1994. Trades are entered when the spread deviates from a k-day moving average, using both calculated over n-days and all available contract months. Positions are closed when the spread returns to its own n-day moving average. Girma and Paulson (1999) suggest to test both 5- and 10-day moving averages and five different entry thresholds. However, a potential problem can occur in cointegration pair trading such as occasionally not finding enough cointegration pairs in the sample (see Chen, Chen, & Chen, 2014.).

Elliott, van der Hoek, and Malcolm (2005) may be the most cited paper using the time-series method on pair trading, as they apply a Kalman filter to estimate a parametric model of the spread and explicitly model the mean reversion behavior of the spread between paired stocks in a continuous time setting. Many studies employ the autoregressive conditional heteroscedastic (ARCH) model by Engle (1982) and generalized ARCH (GARCH) by Bollerslev (1986) for describing dynamic volatility in financial time series. These symmetric GARCH family models have been extended, modified, and refined to handle various types of volatility asymmetry and nonlinearity, e.g. the double threshold GARCH (Chen & So, 2006) and asymmetric smooth transition GARCH (Gerlach & Chen (2008)) models, and many others. Chen et al. (2014) select pairs via the minimum squared distance (MSD) method and construct a pair trading strategy with three-regime threshold autoregressive models with GARCH effects. For the stochastic spread method, Liu and Timmermann (2013) derive optimal portfolio strategies when pairs of risky asset prices are cointegrated so that their conditional excess return can be characterized through a mean reverting error correction process.

In this study we propose to combine the MSD pair selection with suitable trading strategies from nonlinear time-series and quantile forecasting approaches to yield powerful empirical results. In other words, we use the MSD method, which calculates the sum of squared deviations between two normalized stock prices and the chosen minimum value. We also adopt expert opinions to consider some common pair trades of similar companies as suggested by related webpages (e.g. http://www.stockpair.net/guide/pairoptions) that usually operate in the same industrial sector.

Pair trading in practice is typically initiated when the ratio spread is outside two standard deviations of its 12-month historical spread. However, the usual pair trading methods fail to identify potential arbitrage opportunities. In this study, instead of employing the conventional method, we propose to use nonlinear heteroskedastic models to open a pair position, mainly because financial time series often exhitbits certain features – such as clustering and dynamic volatility, asymmetry in conditional mean and variance, mean reversion, and fat-tailed distributions (excess kurtosis). It thus becomes crucial to develop an appropriate model that can capture these stylized facts. This study applies modern versions of such models to the return spreads.

Chen et al. (2014) set up the return spread of potential stock pairs via three-regime threshold autoregressive GARCH (TAR-GARCH) models. While the TAR-GARCH model gives a sharp threshold transition function, a smooth transition model is kind of a continuous mixture between the two underlying regimes and thus has merit to capture a vastly increased and more general set of mean and volatility processes than a sharp transition model. In this study we employ the ST-GARCH models with a second-order logistic function for the return speed. Jansen and Teräsvirta (1996) appear to be the first to examine the second-order logistic function in the ST model. In fact, the ST-GARCH model can be viewed as a model with three regimes, where the first regime is related to extremely low negative shocks; the middle regime represents low absolute returns; and the third regime corresponds to high positive shocks.

This study proposes two methods to trigger buying (entry) and selling (exit) signals to capitalize on market inefficiencies: the first is to use the upper and lower threshold values and the second is to use one-step-ahead quantile forecasts (i.e. Value-at-Risk (VaR) forecasts) obtained from the ST-GARCH model with a second-order logistic function (see Jorion, 1997 for a general review of VaR). We conduct a simulation study and empirical analysis using daily stock returns of 36 U.S. stocks and employ Bayesian Markov chain Monte Carlo (MCMC) sample methods to update the estimates and quantile forecasts. The simulation results strongly support the validity of our sampling scheme to obtain the posterior estimates. We use the minimum square distance method, and the proposed strategies yield average annualized returns of at least 35.5% without a transaction cost and at least 18.4% with a transaction cost.

To our best knowledge, this work is the first to utilize quantile forecasts based on a nonlinear ST-GARCH model for handling an investment strategy. It is well known that quantile forecasting provides more robust results against outliers and model bias than the conventional prediction method, and that trade strategies can be controlled by choosing different quantiles. Moreover, the ST-GARCH model with Students t errors includes a broad class of financial heteroskedastic models and captures very well the characteristics of assets' liquidity. As such, the proposed method provides a functional tool and can be applied to various trading situations in a financial market. From an empirical study, our findings demonstrate that the method is extremely helpful for traders and practitioners and can be implemented efficiently through the Bayesian method, which provides a credible warning signal for when to initiate trades, increase the trading position, or exit the trades.

The remaining part of this study is organized as follows. Section 2 illustrates the ST-GARCH models with second-order logistic functions. Section 3 demonstrates the Bayesian set-up and details of parameter inference. We also describe the process of VaR quantile forecasting. Section 4 presents the pair trading procedure. Section 5 examines the estimation performance of the ST-GARCH model with a second-order logistic function through a simulation study. Section 6 shows empirical results of the daily stock returns of 36 stocks from U.S. stock markets. We further consider pair trading profits for two out-of-sample periods in 2014: two six-month time frames as well as for the entire year. Section 7 provides concluding remarks.

2. The smooth transition heteroskedastic model with a second-order logistic function

Chan and Tong (1986) introduce a smooth transition (ST) model for nonlinear time series, which gained popularity via Granger and Teräsvirta (1993) and Teräsvirta (1994). Ever since then, smooth transition models have been widely used in the modeling of conditional mean or conditional variance - that is, for modeling the asymmetric response of conditional variance to positive and negative news. Regarding the ST function, the literature usually assumes it to be a logistic, exponential, or cumulative distribution. van Dijk, Teräsvirta, and Franses (2002) provide comprehensive reviews of the ST autoregressive model and several of its variants.

This section proposes two statistical methods to find trading entry and exit signals. The first is to fit a ST-GARCH model, with a second-order logistic function and error terms following a standardized Student's t-distribution, to pair return spreads and then to use the estimated upper and lower threshold values in the model as trading entry and exit signals. The second is to utilize one-step-ahead quantile forecasts based on the same model as trading entry and exit signals.

Bollerslev, Chou, and Kroner (1992) advocate that the GARCH (1,1) model is sufficient for most financial time series among general order GARCH models. In the same spirit, to handle the return spread we employ the ST-GARCH(1,1) model with second-order logistic functions as follows:

$$\begin{aligned}
&y_t = \mu_t^{(1)} + F(z_{t-d};\gamma,c_1,c_2)\mu_t^{(2)} + a_t \\
&a_t = \sqrt{h_t}\epsilon_t, \quad \epsilon_t \overset{\text{i.i.d.}}{\sim} t^*(\nu), \\
&h_t = h_t^{(1)} + F(z_{t-d};\gamma,c_1,c_2)h_t^{(2)} \\
&\mu_t^{(i)} = \phi_0^{(i)} + \phi_1^{(i)}y_{t-1} \\
&h_t^{(i)} = \alpha_0^{(i)} + \alpha_1^{(i)}a_{t-1}^2 + \beta_1^{(i)}h_{t-1}, \quad i = 1,2
\end{aligned} \tag{1}$$

$$F(z_{t-d};\gamma,c_1,c_2) = \frac{1}{1+\exp\left\{\frac{-\gamma(z_{t-d}-c_1)(z_{t-d}-c_2)}{s_z}\right\}}, \quad c_1 < c_2, \tag{2}$$

where y_t is the return spread; $h_t = \mathrm{Var}(y_t|\mathscr{F}_{t-1})$; $\mu_t = E(y_t|\mathscr{F}_{t-1})$ with $\mathscr{F}_{t-1}$ representing the information set at time $t-1$; $t^*(\nu)$ is a standardized Student's t-distribution with ν degrees of freedom; z_t is the threshold variable, which can be a past observation of y_t or an exogenous variable; d is the delay lag; s_z is a sample standard deviation of z_t; and $F(.)$ is a continuous distribution changing fully from zero to one. The assumption of shocks normally incurs heavy tails in practice.

Fig. 1 illustrates the second-order logistic ST function for various values of the smoothness parameter γ. We select $c_1 = -1$ and $c_2 = 1$ and notice that when $\gamma = 1$, the shape of $F(\cdot)$ gradually changes to create a smoother, slower transition. When $\gamma = 30$, the transition function starts at 1, decreases to zero within a range of (c_1, c_2), and then returns to 1. It seems that the shapes of $F(\cdot)$ do not exhibit much difference between $\gamma = 10$ and 30. Now we consider the case of $\gamma \to \infty$. If $\gamma \to \infty$and $z_{t-d} < c_1$, then $F(\cdot)$ goes to 1. If $\gamma \to \infty$ and $c_1 < z_{t-d} < c_2$, then $F(\cdot)$ equals 0 during the $c_1 < z_{t-d} < c_2$ interval. Finally, when $\gamma \to \infty$and $z_{t-d} > c_2, F(\cdot)$ goes back to 1 again. Therefore, $F(\cdot)$ can be treated as step function regime transitions and also can be viewed as three regimes, at $z_{t-d} < c_1, c_1 < z_{t-d} < c_2$, and $z_{t-d} > c_2$. For the extreme case $\gamma \to \infty$, Eq. (1) can be viewed as a three-regime threshold model, but the parameters are the same for the upper regime and lower regime.

In order to ensure positive variance and covariance stationarity, we need the following conditions:

$$\begin{aligned}
&\alpha_0^{(1)},\ \alpha_1^{(1)},\ \beta_1^{(1)} > 0,\ \alpha_1^{(1)} + \alpha_1^{(2)} > 0,\ \beta_1^{(1)} + \beta_1^{(2)} > 0, \\
&\quad \left(\alpha_1^{(1)} + 0.5\alpha_1^{(2)}\right) + \left(\beta_1^{(1)} + 0.5\beta_1^{(2)}\right) < 1
\end{aligned} \tag{3}$$

These conditions can also be found in Anderson, Nam, and Vahid (1999). To allow for possible explosive volatility and to ensure a proper prior, Gerlach and Chen (2008) generalize the above two restrictions as follows:

$$\alpha_0^{(1)} < b_1,\ \beta_1^{(1)} < b_2,\ \alpha_1^{(1)} + \beta_1^{(1)} < b_3, \tag{4}$$

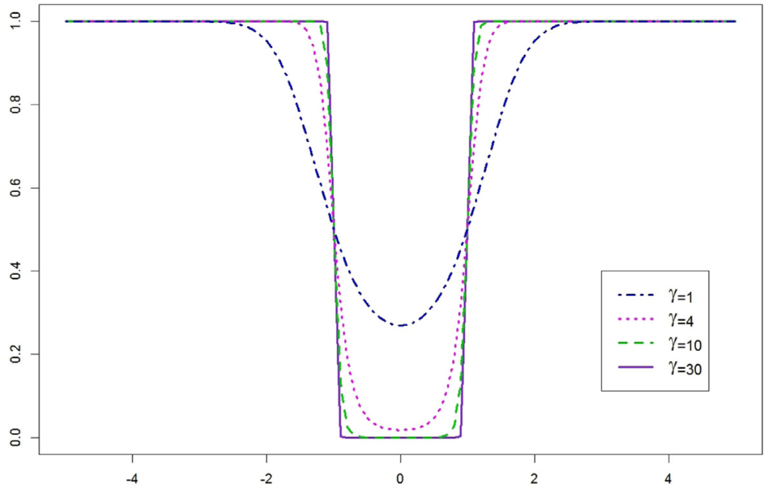

Fig. 1. The shape of the second-order logistic functions for $c_1 = -1$ and $c_2 = 1$.

where b_1, b_2, and b_3 are user-specified, and $b_3 \geqslant 1$ to allow for explosive behavior. We also allow for regime 1 to be possibly explosive.

3. Bayesian inference

Evidence in the literature shows that financial time series models tend to be leptokurtic. Thus, we let ϵ_t follow a standardized Student's t-distribution with v degrees of freedom, and so the conditional likelihood function for the ST model is:

$$p(\boldsymbol{y}^{s+1,n}|\boldsymbol{\theta}) = \prod_{t=s+1}^{n} \left\{ \frac{\Gamma(\frac{v+1}{2})}{\Gamma(\frac{v}{2})\sqrt{(v-2)\pi}} \frac{1}{\sqrt{h_t}} \left[1 + \frac{(y_t - \mu_t)^2}{(v-2)h_t} \right]^{-\frac{v+1}{2}} \right\}, \tag{5}$$

where $\boldsymbol{\theta}$ represents all model parameters; $\boldsymbol{y}^{s+1,n} = (y_{s+1}, \ldots, y_n)$; n is the sample size; and $s = \max\{1, d_0\}$. Note that $\boldsymbol{\theta} = (\boldsymbol{\phi}_i{}', \boldsymbol{\alpha}_i', \boldsymbol{c}', \gamma, v, d)'$ with $\boldsymbol{\phi}_i = (\phi_0^{(i)}, \phi_1^{(i)})'$, and $\boldsymbol{\alpha}_i = (\alpha_0^{(i)}, \alpha_1^{(i)}, \beta_1^{(i)})'$, $i = 1, 2$, and $\boldsymbol{c} = (c_1, c_2)'$.

As $\gamma \to 0$, $F(z_{t-d}; \gamma, c_1, c_2) \to 0.5$, the model reduces to a single regime AR-GARCH, and the parameters in (1) become unidentified. Following the prior set-up in Gerlach and Chen (2008), we design a mixture prior to aid in parameter identification as the smoothing parameter tends to zero, allowing for a sensible inference in that case. We define the latent variable $\delta_j^{(i)}$, which determines the prior distribution of $\phi_j^{(i)}$, via a mixture of two normals:

$$\phi_j^{(i)}|\delta_j^{(i)} \sim (1-\delta_j^{(i)})N(0, k^2 {\tau_j^{(i)}}^2) + \delta_j^{(i)} N(0, {\tau_j^{(i)}}^2), \quad j = 0, 1$$

$$\delta_j^{(i)}|\gamma = \begin{cases} 1, & \text{if } i = 1 \text{ or } \gamma > \xi \\ 0, & \text{if } i = 2 \text{ and } \gamma \leqslant \xi, \end{cases} \tag{6}$$

where ξ is a specified threshold and $\gamma \leqslant \xi$ indicates that $F(z_{t-d}; \gamma, c_1, c_2) \to 0.5$; i.e. a linear AR-GARCH model. As in George and McCulloch (1993), we choose k to be a small positive value so that if $\gamma \leqslant \xi$ and $\delta_j^{(2)} = 0$, then the posterior for the parameters $\phi_j^{(2)}$ will be weighted by the prior towards 0.

Let $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1{}', \boldsymbol{\alpha}_2')'$. We choose a constrained uniform prior on $\boldsymbol{\alpha}$:

$$p(\boldsymbol{\alpha}) \propto \mathrm{I}(S),$$

defined by the indicator $\mathrm{I}(S)$ with S as the set of $\boldsymbol{\alpha}$ that satisfies (3) and (4). Since $\gamma > 0$, we assume γ follows a lognormal distribution: $\ln(\gamma) \sim N(\mu_\gamma, \sigma_\gamma^2)$. This lognormal prior can solve the integrability problem (see Gerlach & Chen, 2008). With maximum delay d_0, we assume the uniform prior $P(d) = \frac{1}{d_0}$, where $d = 1, \ldots, d_0$. For the degrees of freedom v, we reparameterize by defining $v^* = v^{-1}$ and set it to be $I(v^* \in [0, 0.25])$. We choose the priors for the threshold parameters in ST functions as follows:

$$c_1 \sim \mathrm{Unif}(l_1, u_1); \quad c_2|c_1 \sim \mathrm{Unif}(l_2, u_2),$$

where l_1 and u_1 are the respective $\wp_{h1}$ and $\wp_{1-h1-h2}$ percentiles of z_t. Furthermore, we set $u_2 = \wp_{(1-h_2)}$ and $l_2 = c_1 + c^*$, where c^* is a selected number that ensures $c_1 + c^* \leqslant c_2$ and at least $100h_2\%$ of observations are in the range (c_1, c_2). The Bayesian estimates of (c_1, c_2) play an important role for the proposed strategy.

The likelihood in (5) and the priors described above are simply multiplied together to give the conditional posterior kernels.

$$p(\boldsymbol{\theta}_l|\boldsymbol{y}^{s+1,n}, \boldsymbol{\theta}_{\neq l}) \propto p(\boldsymbol{y}^{s+1,n}|\boldsymbol{\theta}) \cdot p(\boldsymbol{\theta}_l|\boldsymbol{\theta}_{\neq l}).$$

Here, $\boldsymbol{\theta}_l$ is each parameter group, $p(\boldsymbol{\theta}_l)$ is its prior density, and $\boldsymbol{\theta}_{\neq l}$ is the vector of all model parameters, except for $\boldsymbol{\theta}_l$. We use the following groups: (i) $\boldsymbol{\phi}_i$, i = 1,2; (ii) $\boldsymbol{\alpha}$; (iii) ν; (iv) γ; (v) d; and (vi) $\boldsymbol{c}$. For all parameters except d, the posterior distributions do not have a standard form. Drawing d is straightforward, noting that it is a multinomial distribution and has the posterior probabilities:

$$p(d = j|\boldsymbol{y}^{s+1,n}, \boldsymbol{\theta}_{\neq d}) = \frac{p(\boldsymbol{y}^{s+1,n}|d = j, \boldsymbol{\theta}_{\neq d})}{\sum_{j=1}^{d_0} p(\boldsymbol{y}^{s+1,n}|d = j, \boldsymbol{\theta}_{\neq d})}, \quad j = 1, \ldots, d_0. \tag{7}$$

Since the conditional posteriors for each of the other parameter groups are non-standard, we incorporate the Metropolis and MH (Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953; Hastings, 1970) methods to draw the MCMC iterates for the groupings $\boldsymbol{\phi}_j, j = 1, 2, \nu, \gamma$, and $\boldsymbol{c}$. For the GARCH parameter $\boldsymbol{\alpha}$, we apply the random walk MH algorithm before the burn-in period and use the independent kernel MH (IK-MH) algorithm after the burn-in period since IK-MH would speed up the convergence.

3.1. Quantile forecasting

The VaR approach is one form of quantile forecasting. Below, we describe how to conduct the one-step-ahead quantile forecasting based on the ST-GARCH model with a second-order logistic function. A one-step-ahead VaR is expressed as the $\alpha\%$ quantile level of the conditional distribution $y_{n+1}|\mathscr{F}_n \sim t(0, h_{n+1})$, where h_{n+1} is the conditional volatility in (1) and t is the Student's t-error distribution. We estimate this predictive distribution via the MCMC simulation. For a one-step-ahead VaR under the smooth transition model in (1) and (2), we can simulate:

$$VaR^{[j]} = -\left[\mu_{n+1}^{[j]} + t_\alpha(\nu^{[j]}) \frac{\sqrt{h_{n+1}^{[j]}}}{\sqrt{\nu^{[j]}/(\nu^{[j]} - 2)}}\right],$$

which forms a posterior predictive MCMC sample of VaR estimates that accounts for parameter uncertainty, where $t_\alpha(\nu^{[j]})$ is the αth quantile of a Student's t-distribution with $\nu^{[j]}$ degrees of freedom; $\sqrt{\nu^{[j]}/(\nu^{[j]} - 2)}$ is an adjustment term for a standardized Student's t with $\nu^{[j]}$ degrees of freedom; and $\mu_{n+1}^{[j]}$, the conditional mean, and $h_{n+1}^{[j]}$, the conditional volatility in (1), $j = M + 1, \cdots, N$, are evaluated conditional upon $\boldsymbol{y}^{s+1,n}$ and the parameter values at MCMC iteration j. The final VaR estimate is the average of this sample. Chen, Weng, and Watanabe (2016) suggest that the ST-GARCH model with a second-order logistic function and skew Students t-error is a worthy choice, when we deal with more extreme VaR forecasts, e.g. 1% level.

4. Pair trading procedure

We choose the trading pairs by using the MSD rule, which finds a corresponding pair that offers the minimum squared distance between the normalized price series for each stock. Since employing the original prices would be a problem for the minimum squared distance rule (because two stocks can move together, but still have a high squared distance between them), we use unit transformation to solve this problem. After normalization, all stocks belong to the same standard unit, which allows trading to be based on a fair formation of pairs. Below, we present our procedure for selecting pairs.

1. Calculate the normalized price p_t^j of asset j at time t:

   $$p_t^j = \frac{P_t^j - E(P^j)}{\sigma^j},$$

   where P_t^j is the closing price of asset $j, E(P^j)$ is the average of P^j, and σ^j is the standard deviation of the respective stock price.
2. Calculate MSD:

   $$MSD = \sum_{t=1}^{n} (p_t^A - p_t^B)^2,$$

where p_t^j is the normalized price of asset j at time t. Pairs are selected by the smallest MSD.

As mentioned earlier, we propose two methods to generate trading entry and exit signals. The first method employs a ST-GARCH model with a second-order logistic function and uses Bayesian estimates of threshold values as trading entry and exit signals, whereas the second method instead uses one-step-ahead quantile forecasts. Here, we also take into consideration the two scenarios of without a transaction cost and with a transaction cost. To use a pair trading strategy, we buy both stocks A and B traded on their first day's price during the out-of-sample period. Below is the pair trading procedure adopted in our study. Among the five steps, Step 2 particularly comprises three different strategies.

Step 1: We calculate the return spread for the in-sample period based on the selected pairs found in the pair selection:

$$y_t = r_t^A - r_t^B, \quad t = 1, \ldots, n.$$

Step 2A: First Strategy: We fit a ST-GARCH model with a second-order logistic function for the return spread and obtain the estimated threshold values by the rolling window approach.

Step 2B: Second Strategy: Based on the ST-GARCH model, we obtain one-step-ahead quantile forecasts for trading signals. We use 20% and 80% quantile forecasts as the lower and upper bounds for trading entry and exit signals.

Step 2C: Third Strategy: This strategy is the same as that of Step 2A except we use 10% and 90% quantile forecasts as the lower and upper bounds for trading entry and exit signals.

Step 3: We check whether we can borrow some treasury shares of stock A at time t. When the return spread is above the "upper bound" at time t, we sell one share of stock A and buy one share of stock B. Conversely, we check whether we can borrow some treasury shares of stock B. If the return spread is below the "lower bound" at time t, then we sell one share of stock B and buy one share of stock A at time t.

Step 4: We calculate the profit gain or loss. If we hold two shares of either one of the two stocks, we use the average price as the cost of this stock for trading purposes. The average trading return is calculated as follows.

1. In case we sell stock A and buy stock B:

$$r_1 = \left[-\ln \frac{P^A_{sold}}{P^A_{bought}} + \ln \frac{P^B_{sold}}{P^B_{bought}} \right].$$

2. In case we sell stock B and buy stock A:

$$r_2 = \left[\ln \frac{P^A_{sold}}{P^A_{bought}} - \ln \frac{P^B_{sold}}{P^B_{bought}} \right].$$

Step 5: If the spread does not cross the threshold levels before the end of the last trading day of the trading period, then we calculate gains or losses at the end of the last trade of the trading period.

We explain the proposed trading entry and exit signals by using a diagram in Fig. 2.

1. We sell one share of stock A and buy one share of stock B at time t_2, which is the first trade.
2. We hold two shares of stock B at time t_2, and so we use the t_1 and t_2 average price as the cost of stock B at time t_1 and t_2.
3. We do not have any share transaction at time t_3, since t_2 and t_3 are in the same upper regime.
4. We do not have any share transaction at time t_4, since t_4 belongs to the middle regime.

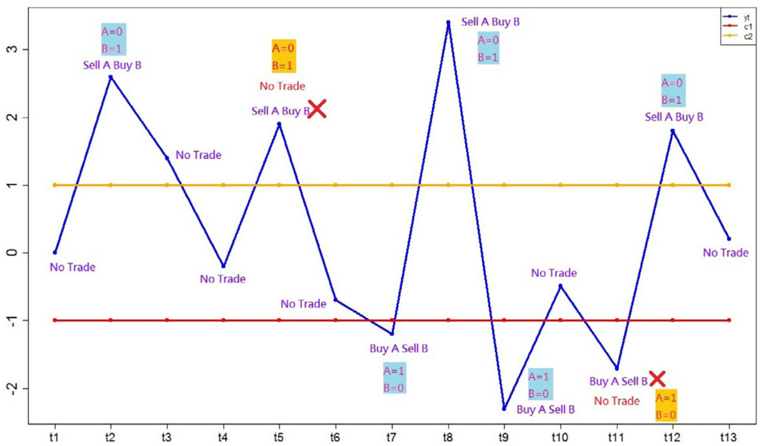

Fig. 2. Diagram for trading entry and exit signals. The red X symbols mean that there are no trade at those time points.

5. We do not have any share transaction at time t_5, since there are no treasury shares of stock A. In this case, we do not trade any stock at this time, which is presented by the first red X symbol in Fig. 2.
6. We buy one share of stock A and sell one share of stock B at time t_7.
7. We have treasury stock, but we do not trade at time t_{11}, which is presented by the second red X symbol in Fig. 2. We only trade when the spread crosses over the upper threshold value at time t_{12}.

5. Simulation study

To examine the effectiveness of the MCMC sampling scheme, we conduct a simulation study. We simulate 500 replicated time series with a sample size of 2000 from the ST-GARCH model:

$$\begin{aligned}
&y_t = (0.1 + 0.4y_{t-1}) + F(z_{t-1})(0.1 - 0.25y_{t-1}) + a_t, \\
&a_t = \sqrt{h_t}\epsilon_t, \epsilon_t \overset{\text{i.i.d.}}{\sim} \sim t^*(7), \\
&h_t = (0.15 + 0.2a_{t-1}^2 + 0.7h_{t-1}) + F(z_{t-1})(-0.1 - 0.1a_{t-1}^2 - 0.2h_{t-1}), \\
&F(z_{t-d}; \gamma, c_1, c_2) = \frac{1}{1+\exp\left\{\frac{-\gamma(z_{t-d}-c_1)(z_{t-d}-c_2)}{s_z}\right\}}, c_1 < c_2,
\end{aligned} \tag{8}$$

where $(\gamma, d) = (5, 1), (c_1, c_2) = (-0.35, 0.3)', z_t$ is the daily returns of the S&P500 index, and s_z is its sample standard deviation.

We use a burn-in sample of $M = 10,000$ and a total sample of $N = 30,000$ iterations, but take only every second iteration in the sample period for inference. For drawing d, we choose the maximum delay $d_0 = 3$. The initial MCMC iterates are $\boldsymbol{\phi}_i = (0, 0)'$, $\boldsymbol{\alpha}_i = (0.1, 0.1, 0.1)', i = 1, 2, \nu = 100, \gamma = 30,$ and $(c_1, c_2) = (0, 0.1)'$. We set the hyper-parameters $(\xi, k) = (0.5, 0.001)$ in the mixture specification (6), $\tau_i = 0.35, (\mu_\gamma, \sigma_\gamma^2) = \left(\ln 5, \left(\frac{\ln 10}{3}\right)^2\right), c_1 \sim \text{Unif}(\wp_{0.2}, \wp_{0.7})$, and $c_2 \sim \text{Unif}(c_1 + c^*, \wp_{0.8})$, where c^* is chosen to be at least 10% of observations in-between c_1 and c_2. Based on the hyper-parameters' set-up of γ, in order for at least 99% of values to lie in the interval we have $(\mu_\gamma - 3\sigma_\gamma, \mu_\gamma + 3\sigma_\gamma)$=$(\ln\frac{1}{2}, \ln 50)$. Hence, the range of γ is about $(e^{\ln\frac{1}{2}}, e^{\ln 50})$=(0.5,50). Following the suggestion by Gerlach and Chen (2008), we set $b_1 = s_y^2$, $b_2 = 1$, and $b_3 = 1.1$, allowing for possible explosive volatility behavior.

In order to confirm convergence and to infer adequate coverage, we extensively examine trace plots and autocorrelation function (ACF) plots from multiple runs of the MCMC sampler for each parameter. The ACF plots appear to cut off fairly quickly, indicating that the MCMC mixing is fast and the autocorrelation is low. To save space, here we do not provide these plots. Table 1 lists the estimation results, wherein the numbers represent the averages of posterior mean, median, standard deviation, and 95% credible interval for 500 replications, except for the parameter d, which is the delay lag. The last row in Table 1 stands for the average of 500 posterior modes.

We now can see that the posterior modes of d correctly indicate $d = 1$ for all 500 datasets. Furthermore, it appears that all estimates are close to their respective true values and the standard deviations for each parameter are small, except for the smoothness parameter γ and $\beta_1^{(1)}$, which is slightly underestimated. Overall, our findings strongly support the validity of our sampling scheme to obtain the posterior estimates.

Table 1
Simulation results for the second-order ST-GARCH model in (8) based on n = 2000 and obtained from 500 replications.

Parameters	True	Mean	Med	Std	2.5%	97.5%
$\phi_0^{(1)}$	0.10	0.1046	0.1068	0.0424	0.0158	0.1821
$\phi_1^{(1)}$	0.40	0.3886	0.3838	0.0916	0.2236	0.5753
$\phi_0^{(2)}$	0.10	0.0961	0.0940	0.0502	0.0041	0.1974
$\phi_1^{(2)}$	−0.25	−0.2403	−0.2366	0.1085	−0.4556	−0.0422
$\alpha_0^{(1)}$	0.15	0.1639	0.1649	0.0339	0.0960	0.2218
$\alpha_1^{(1)}$	0.20	0.2132	0.2062	0.0833	0.0575	0.3891
$\beta_1^{(1)}$	0.70	0.6631	0.6622	0.1219	0.4261	0.8800
$\alpha_0^{(2)}$	−0.10	−0.1035	−0.1045	0.0395	−0.1751	−0.0242
$\alpha_1^{(2)}$	−0.10	−0.1063	−0.0981	0.0968	−0.3055	0.0740
$\beta_1^{(2)}$	−0.20	−0.1751	−0.1827	0.1617	−0.4872	0.1476
ν	7.00	7.1122	6.9896	1.0720	5.3819	9.6005
γ	5.00	5.5563	4.9262	2.5948	2.3668	11.8348
c_1	−0.35	−0.3569	−0.3516	0.1362	−0.6338	−0.1063
c_2	0.30	0.3136	0.3068	0.1392	0.0566	0.6002
d	1	1[a]				

Bayesian estimation: the numbers represent the averages of posterior mean, median, standard deviation, and 95% credible interval for 500 replications, except for the parameter d.
[a] Stands for the average of 500 posterior modes.

Table 2
List of 15 pairs: Pairs 1–10 are selected from MSD criteria and Pairs 11–15 are based on expert opinions for companies that are from the same business sectors.

Pairs	Code	Name	Code	Name	MSD	Same sector
1	DIS	Walt Disney	MMM	3 M	75.4440	
2	DIS	Walt Disney	JNJ	Johnson & Johnson	85.8570	
3	DIS	Walt Disney	TRV	Travelers	99.5785	
4	PG	Procter & Gamble	PEP	PepsiCo	102.0555	
5	CVX	Chevron Corporation	KO	Coca-Cola	118.1018	
6	TRV	Travelers	HD	The Home Depot	121.9505	
7	DIS	Walt Disney	PEP	PepsiCo	129.0492	
8	UTX	United Technologies	MMM	3 M	136.0807	
9	PEP	PepsiCo	JNJ	Johnson & Johnson	140.8767	
10	DIS	Walt Disney	HD	The Home Depot	144.5762	
11	HD	Home Depot	LOW	Lowe's Companies		√
12	GS	Goldman Sachs Group	JPM	JPMorgan Chase		√
13	KO	Coca-Cola	PEP	PepsiCo		√
14	YHOO	Yahoo	GOOGL	Google		√
15	MSFT	Microsoft	AAPL	Apple		√

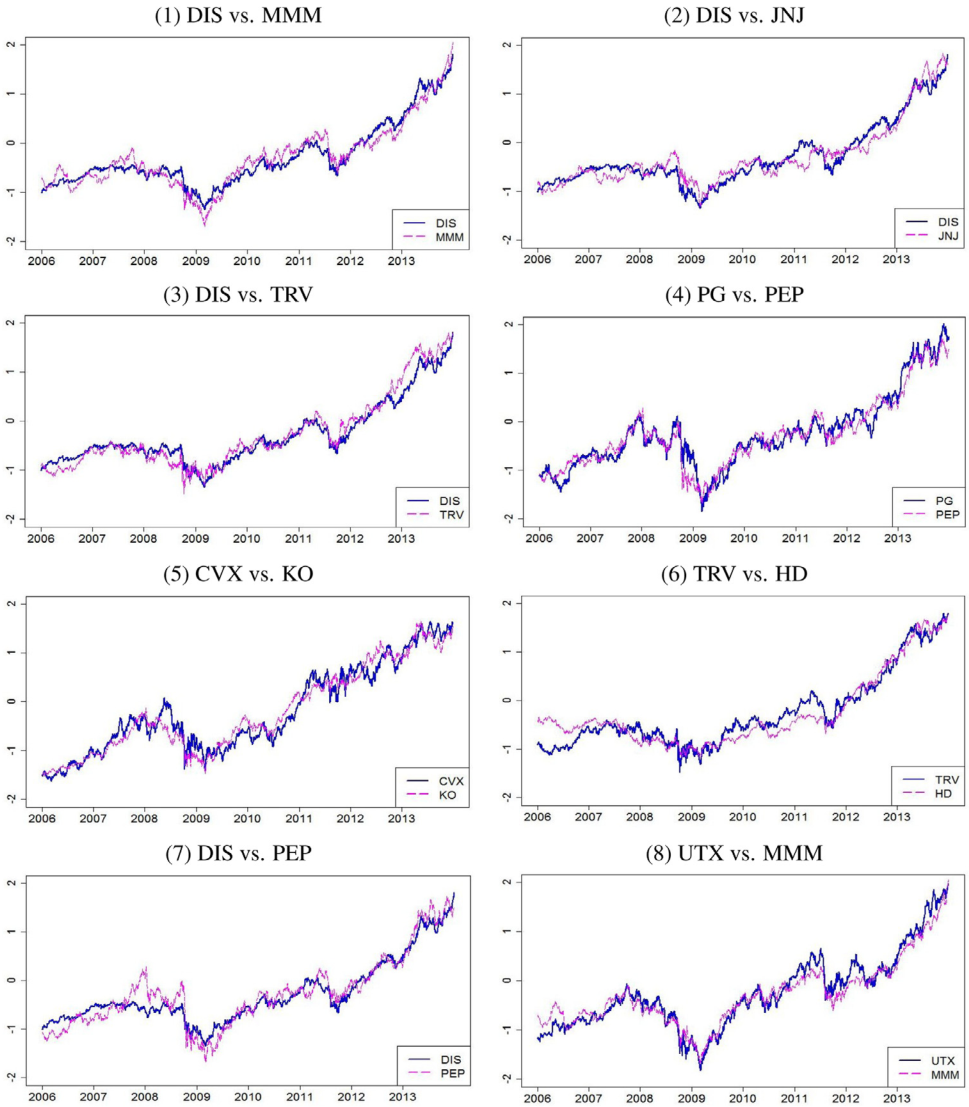

Fig. 3. Time series plot of normalized prices for eight pair trades from January 2006 to December 2013.

6. Empirical study

In selecting pairs, we take 28 stocks from the Dow Jones Industrial Average (DJIA) Index and 8 stocks from the New York Stock Exchange (NYSE) and NASDAQ stock markets. We obtain the data from Yahoo Finance US over a 9-year period, from January 2, 2006 to December 31, 2014. Nike and Visa, also members in the DJIA Index, are excluded here, because they provide fewer datapoints during the in-sample period than the others. The DJIA Index consists of 30 stocks and is broadly recognized as a credible leading index worldwide, because all of them represent companies in important industrial sectors and are blue-chip stocks. The remaining 8 stocks such as Apple and Yahoo are selected, because they are also renowned companies parallel to the components of the DJIA Index. Apple was originally listed on the NASDAQ stock market during the sample period of this study. However, on March 19, 2015, Apple replaced AT&T and is now listed on the Dow Jones Industrial Average.

The in-sample period of this study is from January 2, 2006 to December 31, 2013, i.e. an 8-year period. Since the out-of-sample period is usually about 10% of the sample, for evaluation we pick up the three out-of-sample periods as follows:

(1) January 2, 2014 to June 30, 2014 (named H_1);
(2) July 1, 2014 to December 31, 2014 (named H_2);
(3) January 2, 2014 to December 31, 2014.

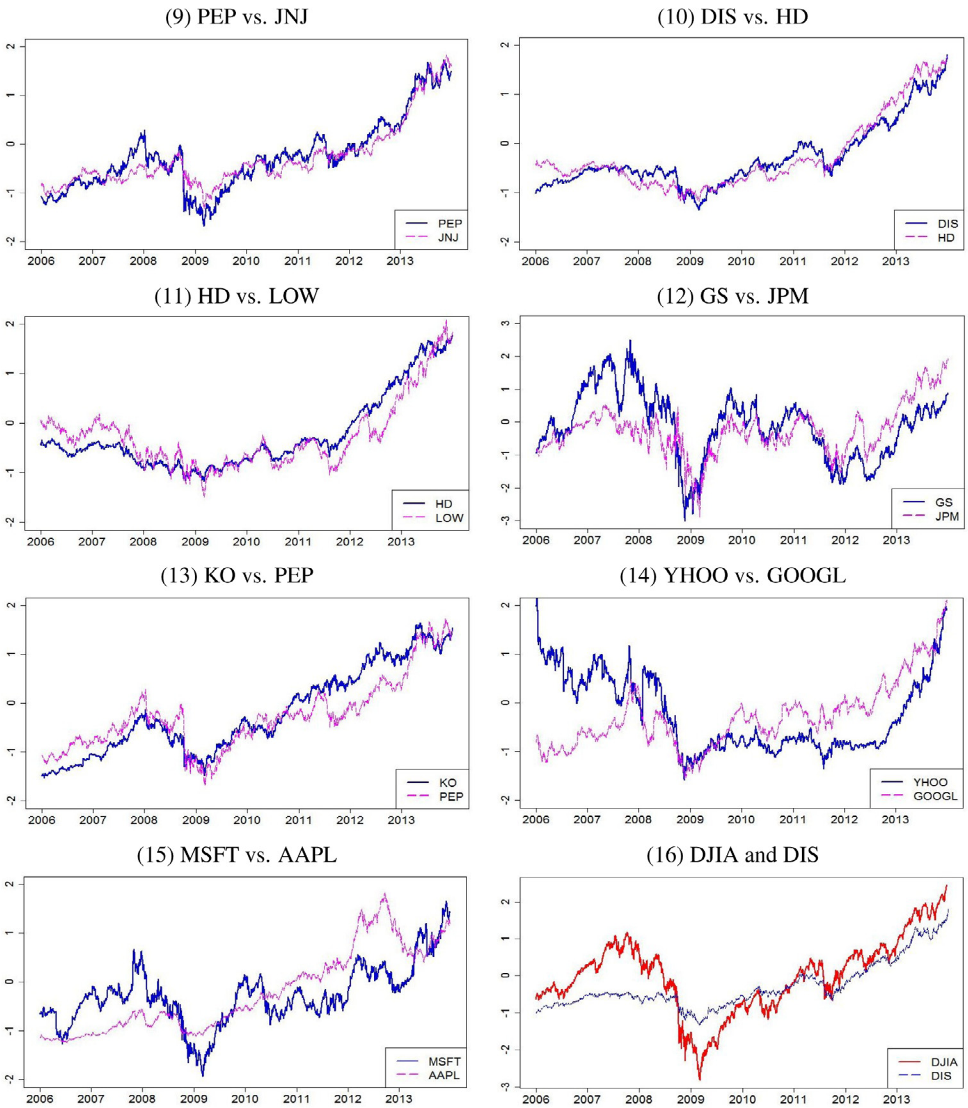

Fig. 4. Time series plot of normalized prices for seven pair trades from January 2006 to December 2013. (16) The normalized prices of DIS vs. DJIA.

Table 3
The descriptive statistics of 36 stock returns during the in-sample period.

Company	Code	Mean	Std.	Minimum	Maximum	Excess Kurtosis	Jarque Bera test
3 M	MMM	0.0401	1.5037	−9.3856	9.4322	1.9392	< 0.001
Alcoa	AA	−0.0435	3.0223	−17.4100	20.8134	2.1214	< 0.001
American Express	AXP	0.0347	2.7011	−19.3791	18.7770	3.6246	< 0.001
Apple	AAPL	0.1039	2.2943	−19.7538	13.0090	5.2684	< 0.001
AT&T	T	0.0388	1.4934	−8.0443	15.1064	1.7946	< 0.001
Bank of American	BAC	0.0424	1.9211	−8.0337	14.3637	2.3882	< 0.001
Boeing	BA	−0.0445	4.0178	−34.1643	30.1018	2.6143	< 0.001
Caterpillar	CAT	0.0321	2.2344	−15.6903	13.7301	2.0375	< 0.001
Chevron Corporation	CVX	0.0164	2.0770	−17.6931	14.7980	2.6799	< 0.001
Cisco Systems	CSCO	0.0526	1.7991	−13.3247	18.9368	3.7387	< 0.001
Coca-Cola	KO	0.0472	1.2227	−9.1044	12.9599	1.928	< 0.001
Dupont	DD	0.0365	1.8964	−12.0474	10.8438	4.1119	< 0.001
ExxonMobil	XOM	0.0386	1.6681	−15.0330	15.8678	2.7427	< 0.001
General Electric	GE	0.0034	2.1228	−13.6859	17.9608	3.5587	< 0.001
Google	GOOGL	0.0494	2.0285	−12.3425	18.2255	2.047	< 0.001
Goldman Sachs Group	GS	0.0206	2.7468	−21.0267	23.4854	1.6544	< 0.001
Hewlett–Packard	HPQL	0.0041	2.1880	−22.351	15.7700	2.5623	< 0.001
IBM	IBM	0.0481	1.4294	−8.6400	10.8968	5.0425	< 0.001
Intel	INTC	0.0142	1.9737	−13.266	11.2253	4.1042	< 0.001
Johnson & Johnson	JNJ	0.0332	1.0356	−7.9895	11.5533	1.9239	< 0.001
JPMorgan Chase	JPM	0.0291	3.0019	−23.2210	22.3767	4.5891	< 0.001
Lowe's Companies	LOW	0.0259	2.1068	−10.8976	15.2724	1.7416	< 0.001
McDonalds	MCD	0.0644	1.2547	−8.3201	8.9829	4.0147	< 0.001
Merck	MRK	0.0391	1.7356	−15.9590	11.9183	2.3167	< 0.001
Microsoft	MSFT	0.0264	1.8317	−12.4280	17.0958	3.7025	< 0.001
PepsiCo	PEP	0.0276	1.1534	−12.7106	8.2200	3.3502	< 0.001
Pfizer	PFE	0.0315	1.5188	−11.2514	9.6863	3.3293	< 0.001
Procter & Gamble	PG	0.0280	1.1690	−8.2143	9.7254	3.5798	< 0.001
The Home Depot	HD	0.0462	1.8525	−8.5844	13.1723	2.1005	< 0.001
Travelers	TRV	0.0453	2.0366	−20.0522	22.7529	3.379	< 0.001
UnitedHealth Group Incorporated	UNH	0.0125	2.3459	−20.6293	29.8633	2.3481	< 0.001
United Technologies Corporation	UTX	0.0444	1.6272	−9.1788	12.7863	3.3147	< 0.001
Verizon Communtions	VZ	0.0497	1.4585	−8.4341	13.6791	2.6753	< 0.001
Wal-Mart	WMT	0.0343	1.2572	−8.4244	10.4823	2.7921	< 0.001
Walt Disney	DIS	0.0640	1.8538	−10.2114	14.8206	2.0561	< 0.001
Yahoo	YHOO	0.0016	2.6524	−24.6364	39.1817	2.3092	< 0.001

All stock returns display positive excess kurtosis, i.e. heavy-tailed distributions. The normality assumption is turned down in all cases at the 1% significance level by the Jarque–Bera normality test.

Table 4
The descriptive statistics of 15 pairs' return spreads during the in-sample period.

	Company A	Company B	Mean	Std.	Minimum	Maximum	Excess Kurtosis	Jarque Bera test
1	DIS	MMM	0.0239	1.4428	−9.0833	13.2028	12.1433	< 0.001
2	DIS	JNJ	0.0308	1.5142	−7.1354	11.3847	7.4571	< 0.001
3	DIS	TRV	0.0187	1.7856	−19.5526	14.1717	16.1890	< 0.001
4	PG	PEP	0.0004	1.0582	−5.2427	13.9783	20.1892	< 0.001
5	CVX	KO	0.0054	1.5459	−14.4224	13.1273	11.8956	< 0.001
6	TRV	HD	−0.0009	1.8886	−14.3231	19.2028	14.9882	< 0.001
7	DIS	PEP	0.0364	1.5626	−8.1316	11.1194	7.6265	< 0.001
8	UTX	MMM	0.0043	1.1904	−6.7779	8.2040	9.8150	< 0.001
9	PEP	JNJ	−0.0056	1.0208	−14.7849	5.8856	27.3897	< 0.001
10	DIS	HD	0.0178	1.6617	−8.3870	11.6734	7.1670	< 0.001
11	HD	LOW	0.0204	1.2372	−6.7935	11.8655	9.8635	< 0.001
12	GS	JPM	−0.0085	2.0929	−25.9269	21.4998	28.7847	< 0.001
13	KO	PEP	0.0195	0.9642	−5.5857	7.0618	9.0484	< 0.001
4	YHOO	GOOGL	−0.0478	2.6604	−23.6250	48.1474	63.8228	< 0.001
15	MSFT	AAPL	−0.0775	2.2202	−18.8205	13.2528	10.4595	< 0.001

All spreads exhibit positive excess kurtosis, i.e. heavy-tailed distributions. The normality assumption is turned down in all cases at the 1% significance level by the Jarque–Bera normality test.

We calculate all MSDs between any two normalized price series: the number of possible pairs is 630 (i.e. C_2^{36}) combinations. We select ten pairs via the MSD method and another five pairs are from two companies in the same industrial sector. Table 2 lists these 15 pairs, which include ten of the smallest MSD values. Note that the first ten pairs are chosen by the MSD method, while pairs 11–15 are two companies in the same industrial sector. From Figs. 3,4 we see the time plots of normalized prices for these 15 pairs and show that the patterns of the first ten pairs chosen by the MSD are quite close to each other. From Fig. 4 (11)-(15) we see 5 pairs chosen by some common pair trade strategies on financial webpages, showing that most trends of the pairs are similar. The patterns of two pairs (YHOO vs. GOOGL and MSFT vs. AAPL), which are from the same business sectors, are not very analogous as seen in Fig. 4.

In the ten pairs with the smallest MSD values, Walt Disney Company (DIS) is selected five times, which suggests that the pattern of DIS is similar to that of DJIA. Fig. 4 (16) presents the normalized prices of DIS and DJIA during the in-sample period. Tables 3 and 4 contain some descriptive statistics and the Jarque–Bera normality test for each stock return and spread. In terms of a stock return's standard deviation, Johnson & Johnson (JNJ) has the smallest value, while Boeing (BA) has the largest. All stock returns and spreads display positive excess kurtosis, i.e. heavy-tailed distributions. In fact, the normality assumption is turned down in all cases at the 1% significance level by the Jarque–Bera normality test, providing strong evidence that ϵ_t in Eq. (1) should be fat-tailed.

We next conduct the three strategies as described in Section 4 based on the ST-GARCH model in Eq. (1) to find the trading entry and exit signals. We select the threshold variable z_{t-d} as the past value of spreads with a maximum delay $d_0 = 3$. In the

Table 5
Bayesian estimation of parameters for the second-order ST-GARCH specifications for the spread of (TRV, HD).

Parameters	Mean	Md	Std.	2.5%	97.5%
$\phi_0^{(1)}$	−0.3016	−0.2973	0.1290	−0.5604	−0.0532
$\phi_1^{(1)}$	0.4188	0.4214	0.2081	0.0142	0.8438
$\phi_0^{(2)}$	0.3815	0.3786	0.1648	0.0601	0.7071
$\phi_1^{(2)}$	−0.4522	−0.4564	0.2162	−0.8905	−0.0334
$\alpha_0^{(1)}$	0.1890	0.1897	0.0506	0.0872	0.2797
$\alpha_1^{(1)}$	0.0822	0.0753	0.0419	0.0069	0.1681
$\beta_1^{(1)}$	0.7208	0.7213	0.0493	0.6280	0.8051
$\alpha_0^{(2)}$	−0.1417	−0.1450	0.0663	−0.2444	−0.0129
$\alpha_1^{(2)}$	−0.0404	−0.0347	0.0422	−0.1254	0.0363
$\beta_1^{(2)}$	0.2672	0.2643	0.0573	0.1702	0.3744
ν	5.2246	5.1699	0.5705	4.2794	6.5019
γ	2.8465	2.5853	1.2758	1.3290	5.7044
c_1	0.0866	0.1168	0.2106	−0.4118	0.4105
c_2	0.7064	0.7412	0.1469	0.3283	0.8813
d	1[a]				

All parameters are significant from 0 except for c_1 for this pair. These estimates perform differently when the rolling window span is allowed to move. $\hat{\nu} \doteq 5$, suggesting the existence of conditional leptokurtosis and justifying the usage of a fat-tailed error distribution.
[a] Pr(d = 1) = 1.

Table 6
Company monthly returns from January 2, 2014 to December 31, 2014.

Pairs	Company A	H1 %	H2 %	Annual %	Company B	H1 %	H2 %	Annual %
1	DIS	1.8598	1.9124	1.7617	MMM	0.5443	2.5879	1.4472
2	DIS	1.8598	1.9124	1.7617	JNJ	2.3721	0.2206	1.2688
3	DIS	1.8598	1.9124	1.7617	TRV	0.807	2.0187	1.4217
4	PG	−0.3180	2.6792	1.1388	PEP	1.4326	1.389	1.2638
5	CVX	0.9919	−1.994	−0.5751	KO	0.6502	0.3984	0.4069
6	TRV	0.8070	2.0187	1.4217	HD	−0.0863	4.0046	2.0966
7	DIS	1.8598	1.9124	1.7617	PEP	1.4326	1.6334	1.2638
8	UTX	0.3976	0.3683	0.2524	MMM	0.5443	2.5879	1.4472
9	PEP	1.4326	1.6334	1.2638	JNJ	2.3721	0.2206	1.2688
10	DIS	1.8598	1.9124	1.7617	HD	−0.0863	4.2115	2.0966
11	HD	−0.0863	4.0046	2.0966	LOW	−0.3934	5.7604	2.7374
12	GS	−0.8361	2.6354	0.8129	JPM	0.0369	1.5757	0.7476
13	KO	0.6502	0.3984	0.4069	PEP	1.4326	1.3890	1.2638
14	YHOO	−2.2704	5.7700	1.7647	GOOGL	0.6689	−1.3408	−0.4401
15	MSFT	1.9856	2.2496	1.9328	AAPL	2.5738	3.3787	2.7060

Regarding individual stocks, AAPL has the maximum monthly average return (2.57%) in the first half of 2014, while YHOO has the minimum monthly return (−2.27%) for the same period.

out-of-sample period, we employ a rolling window approach to produce estimated threshold values and one-step-ahead quantile forecasting of (20%, 80%) and (10%, 90%) over 15 pairs. We perform 30,000 MCMC iterations and discard the first 10,000 iterates as a burn-in sample for each spread during each moving window. The setting for initial MCMC iterates is the same as in the simulation study.

To save space, Table 5 only lists posterior summaries for the pair trade TRV vs. HD, including posterior median, standard deviation, and the 95% credible interval of the unknown parameters. Most of the parameters appear to be significant, as indicated by the 95% credible intervals. The estimate of c_1 is not significantly different from zero for this pair. However, these estimates perform differently when the rolling window span is allowed to move. Regarding the error distribution of ϵ_t, all degrees of freedom estimates are less than 10, suggesting the existence of conditional leptokurtosis and justifying the usage of a fat-tailed error distribution. The results are available from the authors upon request.

6.1. Strategy profits

As seen in Table 6, AAPL has the maximum monthly average return (2.57%) in the first half of 2014, while YHOO has the minimum monthly return (−2.27%) for the same period. Based on Strategies 1–3, we list the results of the first half year in

Table 7
The pair trades' semi-annual profits based on Strategies 1–3 for the first half of 2014. In this out-of-sample period, we employ a rolling window approach to produce estimated threshold values for Strategy 1 and one-step-ahead quantile forecasting of (20%, 80%) and (10%, 90%) for Strategies 2–3 over 15 pairs.

			Strategy 1		Strategy 2		Strategy 3	
	Company A	Company B	Round-trip trades	Pair Return %	Round-trip trades	Pair Return %	Round-trip trades	Pair Return %
1	DIS	MMM	21	13.2706	11	17.8360	5	24.6459
2	DIS	JNJ	20	25.8145	9	26.1914	6	23.2576
3	DIS	TRV	20	14.3095	10	19.1258	10	19.1258
4	PG	PEP	22	7.7760	12	6.2913	4	2.4276
5	CVX	KO	23	12.3796	14	12.1063	10	15.1102
6	TRV	HD	22	6.1961	12	7.2715	6	7.8719
7	DIS	PEP	23	22.6500	11	18.5872	10	23.5176
8	UTX	MMM	21	8.5186	14	7.9637	5	9.9440
9	PEP	JNJ	20	24.9391	9	23.8647	3	23.8647
10	DIS	HD	21	10.1090	10	10.3438	5	16.0878
11	HD	LOW	27	−5.0058	13	−1.2040	4	−4.5453
12	GS	JPM	14	−7.1316	8	−4.4227	4	−10.2988
13	KO	PEP	23	17.8086	9	13.6526	5	8.7746
14	YHOO	GOOGL	16	−11.3828	12	−7.1254	6	−12.1003
15	MSFT	AAPL	22	30.3414	14	31.0696	5	37.0520

The profits of the first ten pairs calculated by Strategies 1–3 are positive. The maximum profit is 30.34% by Strategy 1 for the pair of MSFT vs. AAPL. Based on Strategy 2, MSFT vs. AAPL has the highest profit (31.07%). Strategy 3 has the largest average semi-annual profit and the lowest number of round-trip trades among the three strategies during the H_1 period.

Table 8
The pair trades' semi-annual profits based on Strategies 1–3 for the second half of 2014. In this out-of-sample period, we employ a rolling window approach to produce estimated threshold values for Strategy 1 and one-step-ahead quantile forecasting of (20%, 80%) and (10%, 90%) for Strategies 2–3 over 15 pairs.

			Strategy 1		Strategy 2		Strategy 3	
	Company A	Company B	Round-trip trades	Pair Return %	Round-trip trades	Pair Return %	Round-trip trades	Pair Return %
1	DIS	MMM	24	28.0772	13	20.8992	4	19.1284
2	DIS	JNJ	23	10.1910	12	12.0447	5	8.7458
3	DIS	TRV	22	23.9380	9	25.7754	2	19.5137
4	PG	PEP	20	21.1917	10	24.1219	4	20.5320
5	CVX	KO	14	−6.9078	9	−11.8894	4	−17.0963
6	TRV	HD	25	38.9275	13	35.3974	4	38.5220
7	DIS	PEP	26	14.2931	12	13.2403	7	13.2998
8	UTX	MMM	22	16.2027	8	20.3759	5	15.0919
9	PEP	JNJ	21	10.9817	11	3.7635	4	9.2260
10	DIS	HD	20	33.3813	11	31.8568	6	36.1855
11	HD	LOW	21	58.5577	7	63.0565	1	45.4682
12	GS	JPM	16	23.9321	8	30.2647	3	16.4170
13	KO	PEP	20	13.6043	10	10.9826	6	7.4188
14	YHOO	GOOGL	23	27.1388	10	25.8844	4	33.2401
15	MSFT	AAPL	24	33.0581	9	27.1607	5	31.6145

All of the pairs show positive profits except for the fifth pair of CVX vs. KO. All three strategies' highest profit occurs for the pair of HD vs. LOW.

Table 7, which includes the returns from a pair trading strategy. MSD ranks the top ten pairs, and we set up a pair trade for those within a given rank. In Table 7, the profits of the first ten pairs calculated by Strategies 1–3 are positive. The maximum profit is 30.34% by Strategy 1 for the pair of MSFT vs. AAPL. Based on Strategy 2, MSFT vs. AAPL has the highest profit (31.07%). Strategy 3 has the largest average semi-annual profit and the lowest number of round-trip trades among the three strategies during H_1 period.

Table 8 lists the pair trades' semi-annual profits during the second half of 2014. Table 8 shows that in the three strategies, all of the pairs exhibit positive profits except for the fifth pair of CVX vs. KO. All three strategies' highest profit occurs with the pair of HD vs. LOW. Table 9 shows the profit for the entire year of 2014 under the three strategies. Based on Strategy 1,

Table 9
The pair trades' annual profits based on Strategies 1–3 for the entire year of 2014. In this out-of-sample period, we employ a rolling window approach to produce estimated threshold values for Strategy 1 and one-step-ahead quantile forecasting of (20%, 80%) and (10%, 90%) for Strategies 2–3 over 15 pairs.

			Strategy 1		Strategy 2		Strategy 3	
	Company A	Company B	Round-trip trades	Pair Return %	Round-trip trades	Pair Return %	Round-trip trades	Pair Return %
1	DIS	MMM	46	43.5887	20	38.7979	10	51.4138
2	DIS	JNJ	45	40.6765	20	42.8567	11	34.8991
3	DIS	TRV	43	44.0021	19	43.1811	19	43.1810
4	PG	PEP	45	35.1969	23	32.5135	8	35.6756
5	CVX	KO	49	2.5591	27	0.2533	21	−1.1797
6	TRV	HD	46	46.6606	25	38.8860	10	39.5239
7	DIS	PEP	48	33.5847	24	37.7517	23	41.4402
8	UTX	MMM	42	27.1425	24	27.2739	10	26.3094
9	PEP	JNJ	39	32.6499	20	31.4784	9	31.4808
10	DIS	HD	41	50.7572	22	44.1789	10	44.8115
11	HD	LOW	50	67.2558	18	61.9223	5	55.9944
12	GS	JPM	35	17.5860	16	14.7646	7	28.8614
13	KO	PEP	44	20.7379	19	28.3349	12	15.7730
14	YHOO	GOOGL	38	18.3189	21	20.9450	10	19.4388
15	MSFT	AAPL	47	62.5986	25	69.6590	10	68.0691

Based on Strategy 1, the pair of HD vs. LOW has the highest profit (67.26%), while for both Strategies 2 and 3, MSFT vs. AAPL has the highest profit of (69.66%) and (68.07%), respectively. All of the pairs exhibit a positive profit for Strategies 1–3, except for the fifth pair (CVX vs. KO) under Strategy 3.

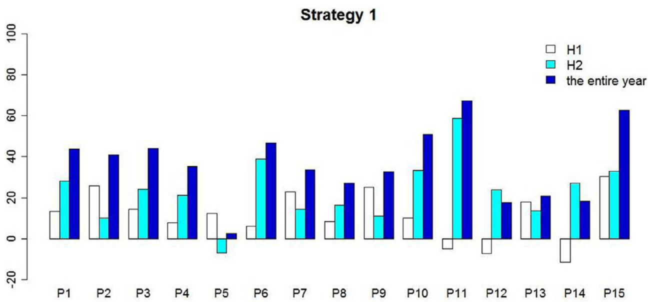

Fig. 5. Returns for 15 pair trades using Strategy 1.

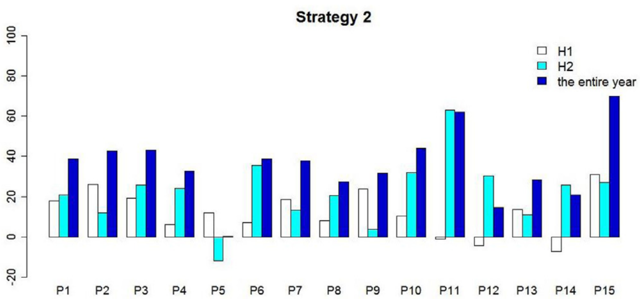

Fig. 6. Returns for 15 pair trades using Strategy 2.

the pair of HD vs. LOW has the highest profit (67.26%), while for both Strategies 2 and 3, MSFT vs. AAPL has the highest profit of (69.66%) and (68.07%), respectively. All of the pairs exhibit a positive profit for Strategies 1–3, except for the fifth pair (CVX vs. KO) under Strategy 3. Fortunately, the loss is only −1.18%.

Tables 7–9 show that the most profitable pairs are those frequently traded with Strategy 1. Figs. 5–7 present the profits of Strategies 1–3 for the three time periods. In Fig. 5 most of the profits from pair trading for the entire year are higher than for both the first half and second half of the year, and three of the pairs even have negative profits in the first half of the year. Fig. 6 shows that the top-ranked minimum profit pairs are similar to those found with Strategy 1 in Fig. 5, and Pair 5 (CVX vs. KO) also has a negative profit in the second half of the year. Fig. 7 lists the top-three profit pairs as Pair 15 (MSFT vs. AAPL) (68.07%), Pair 11 (HD vs. LOW) (55.99%), and Pair 1 (DIS vs. MMM) (51.41%). We notice that two out of the three highest

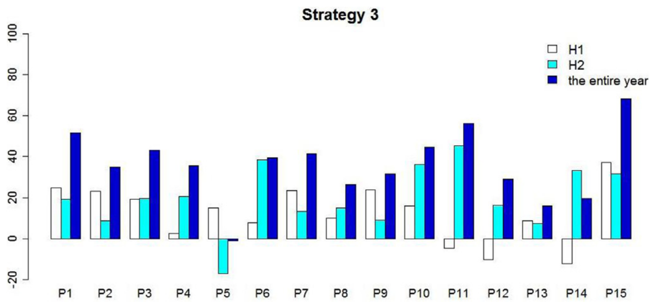

Fig. 7. Returns for 15 pair trades using Strategy 3.

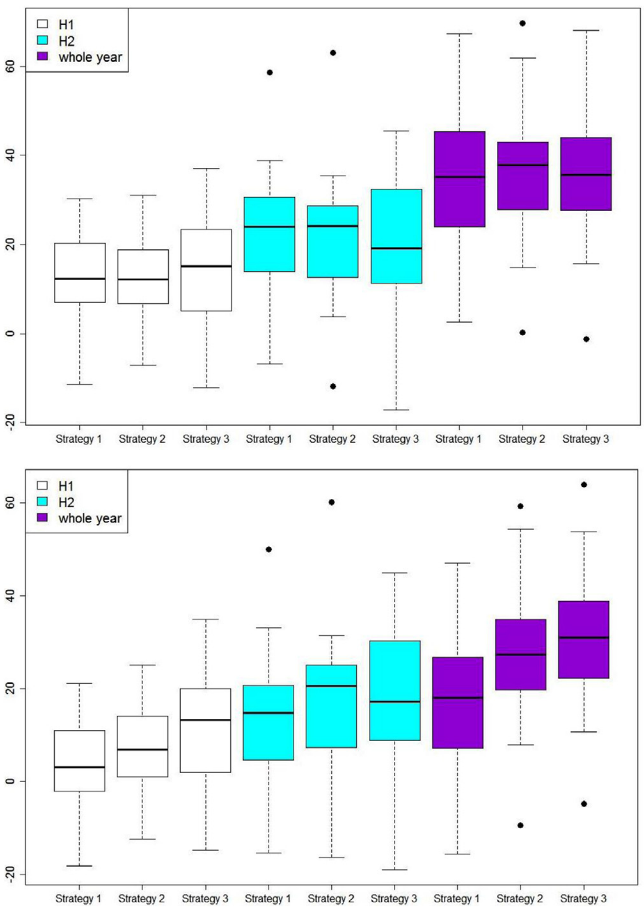

Fig. 8. Return profits for the three strategies. Upper panel: without a transaction cost. Lower panel: with a 0.2% transaction cost.

annualized returns come from the same business sector. All these results demonstrate that the proposed strategies provide an excellent timing for pair trading.

Fig. 8 displays the boxplot of profits for the three strategies during the three out-of-sample periods. As shown in the upper panel of Fig. 8, the range of Strategy 1 is the widest among the three strategies for the entire year. Fig. 9 presents the esti-

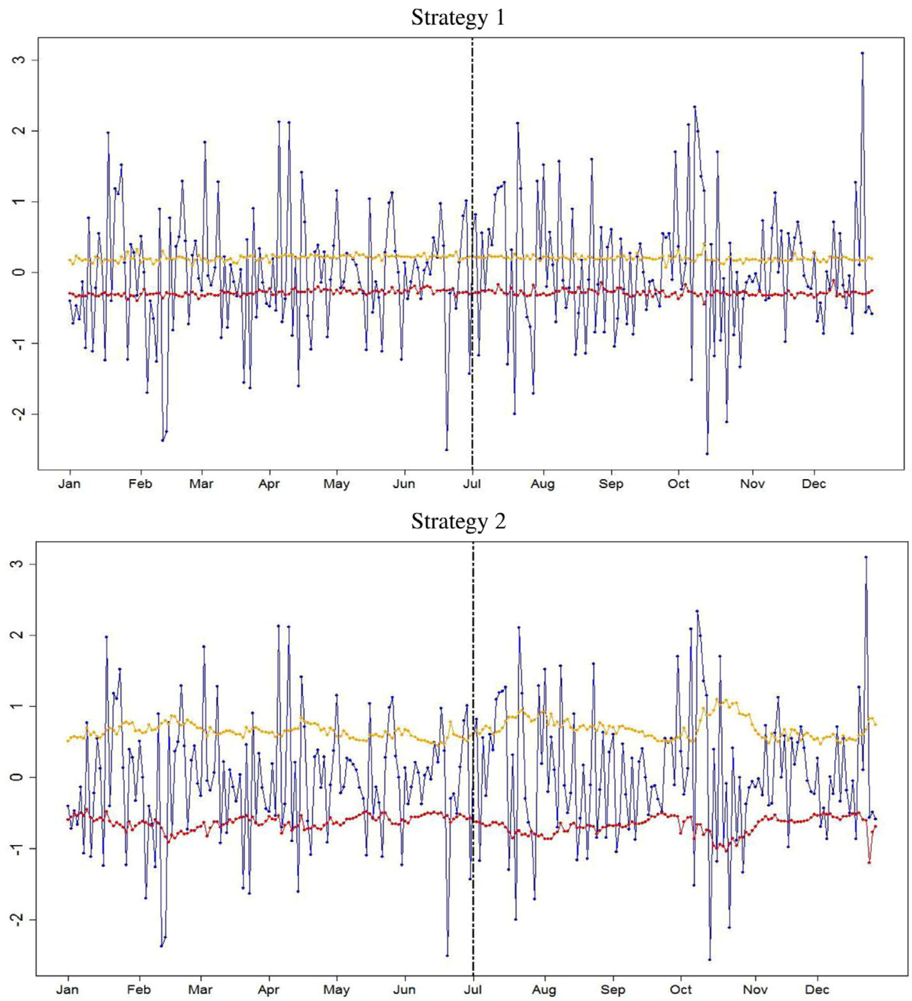

Fig. 9. Daily trading entry and exit signals for PEP/JNJ based on Strategies 1 and 2.

Table 10
Summary statistics of round-trip trades. For the three out-of sample periods, Strategy 1 has the largest number of round-trip trades. The most profitable pairs are those frequently traded with Strategy 1, whereas Strategy 3 is more conservative in terms of number of trades than Strategies 1 and 2.

	Mean	Median	Std.	Minimum	Maximum
H_1					
Strategy 1	21.0	21	3.0	14	27
Strategy 2	11.2	11	2.0	8	14
Strategy 3	5.9	5	2.3	3	10
H_2					
Strategy 1	21.4	22	3.2	14	26
Strategy 2	10.1	10	1.8	7	13
Strategy 3	4.3	4	1.5	1	7
Annual					
Strategy 1	43.9	45	4.2	35	50
Strategy 2	21.5	21	3.1	16	27
Strategy 3	11.7	10	5.2	5	23

mated daily trading entry and exit signals for Strategies 1 and 2 that employ the rolling window approach for the pair PEP vs. JNJ. It shows that the pattern of quantile forecasts for Strategies 1 and 2 looks similar, and thus, we only provide quantile forecasts for Strategy 2. The trading signals of Strategy 1 are much narrower than those of Strategies 2 and 3, indicating that Strategy 1 experiences more frequent trading.

Table 10 summarizes the number of round-trip trades. For the three learning periods, Strategy 1 has the largest number of round-trip trades. Moreover, it is shown that the most profitable pairs are those frequently traded with Strategy 1, whereas Strategy 3 is more conservative in terms of number of trades than Strategies 1 and 2.

Table 11
Profits for the three strategies with a 0.2% transaction cost for the entire year of 2014. We assume 0.2% cost to go short and nothing to go long.

	Company A	Company B	Strategy 1		Strategy 2		Strategy 3	
			Round-trip trades	Pair Return %	Round-trip trades	Pair Return %	Round-trip trades	Pair Return %
1	DIS	MMM	46	24.7887	20	30.5979	10	47.2138
2	DIS	JNJ	45	22.4765	20	34.6567	11	30.0991
3	DIS	TRV	43	26.6021	19	35.1811	19	35.1810
4	PG	PEP	45	−15.6409	23	−9.3467	8	−4.7797
5	CVX	KO	49	26.8606	27	27.8860	21	30.9239
6	TRV	HD	46	14.9847	25	27.3517	10	37.0402
7	DIS	PEP	48	13.2499	24	21.6784	23	22.0808
8	UTX	MMM	42	17.9969	24	22.5135	10	31.2756
9	PEP	JNJ	39	11.1425	20	19.0739	9	22.5094
10	DIS	HD	41	33.9572	22	35.1789	10	40.6115
11	HD	LOW	50	47.0558	18	54.3223	5	53.7944
12	GS	JPM	35	3.3860	16	7.9646	7	25.8614
13	KO	PEP	44	2.7379	19	20.3349	12	10.7730
14	YHOO	GOOGL	38	2.7189	21	12.1450	10	15.0388
15	MSFT	AAPL	47	43.5986	25	59.2590	10	63.8691

Pairs 15 (MSFT vs. AAPL), 11 (HD vs. LOW), and 10 (DSI vs. HD) present the three largest annual profits for Strategies 1 and 2, while pairs 15, 11, and 1 (DIS vs. MMM) show the three highest annual profits for Strategy 3. The annualized returns for the top three pairs from the three strategies are at least 33.95% with transaction costs.

Table 12
Summary of pair trading profits for the three strategies without a transaction cost and with a 0.2% transaction cost. We assume 0.2% cost to go short and nothing to go long.

	Mean	Median	Std.	Minimum	Maximum
Without a transaction cost					
H_1					
Strategy 1	11.3729	12.3796	12.2830	−11.3828	30.3414
Strategy 2	12.1034	12.1063	11.0855	−7.1254	31.0696
Strategy 3	12.3157	15.1102	13.9859	−12.1003	37.0520
H_2					
Strategy 1	23.1045	23.9321	15.0902	−6.9078	58.5577
Strategy 2	22.1956	24.1219	16.6318	−11.8894	63.0565
Strategy 3	19.8205	19.1284	15.6379	−17.0963	45.4682
Annual					
Strategy 1	36.2210	35.1969	17.4390	2.5591	67.2558
Strategy 2	35.5198	37.7517	17.1320	0.2533	69.6590
Strategy 3	35.7128	35.6756	17.0464	−1.1797	68.0691
With a transaction cost					
H_1					
Strategy 1	2.6529	3.1796	11.9118	−18.1828	30.3414
Strategy 2	7.2901	6.9063	11.0780	−12.3254	31.0696
Strategy 3	9.7024	13.3102	13.7986	−14.7003	37.0520
H_2					
Strategy 1	14.2512	14.7380	15.4990	−15.3078	58.5577
Strategy 2	17.8490	20.5219	17.0848	−16.2894	63.0565
Strategy 3	17.8072	17.1284	15.8177	−18.8963	45.4682
Annual					
Strategy 1	18.3944	17.9969	16.6065	−15.6409	67.2558
Strategy 2	26.5865	27.3517	17.0098	−9.3467	69.6590
Strategy 3	30.7662	30.9239	17.1785	−4.7797	68.0691

Without considering transaction cost, Strategy 1 has the largest average semi-annual profit among the three strategies during H_2 and has the largest average annual profit during the entire year. When considering the 0.2% transaction cost, Strategy 3 appears to be the most profitable strategy among the three.

We now turn to consider the performance of profits for the three strategies with 0.2% transaction cost. In other words, we assume an upfront 0.2% cost to go short, while there is no cost for going long. To save space, we only present profits for the three strategies with this 0.2% transaction cost for the entire year in Table 11 as well as the boxplot of profits for the three strategies during the three out-of-sample periods in Fig. 8. Pairs 15 (MSFT vs. AAPL), 11 (HD vs. LOW), and 10 (DSI vs. HD) present the three largest annual profits for Strategies 1 and 2, while pairs 15, 11, and 1 (DIS vs. MMM) show the three highest annual profits for Strategy 3. The annualized returns for the top three pairs from the three strategies are at least 33.95% with this transaction cost.

Table 12 presents the summary statistics of the trade-return profits for the 15 pairs with and without a transaction cost. Without considering any transaction cost, the cumulative mean net profit from the first half of the year is 11.37%, 12.10%, and 12.32% for Strategies 1–3, respectively. For the second half of the year, it is 23.10%, 22.20%, and 19.82% for Strategies 1–3, respectively. Without considering any transaction cost, Strategy 1 has the largest average semi-annual profit among the three strategies during H_2 and has the largest average annual profit during the entire year. However, the result becomes very different when considering the 0.2% transaction cost. In this case, Strategy 3 appears to be the most profitable strategy among the three.

7. Concluding remarks

Pair trading is a mean-reverting statistical arbitrage technique based on the assumption that prices should revert to their historical trends. In this study, after setting up pairs of stocks, we select those pairs with a minimum distance between their normalized historical prices and then propose three statistical arbitrage strategies implemented based on the trading entry and exit signals generated from the fitted ST-GARCH model to the return spreads. Our trading strategies particularly have merit for being independent of market movements - in other words, they are market neutral.

In this study we systematically have examined the comparative predictive performance of three trading strategies for stock returns through an empirical study. As for trading entry and exit signals, we propose to use either the upper and lower threshold values of ST-GARCH models or one-step-ahead quantile forecasts based on the same model. In this procedure, we use Bayesian MCMC sampling scheme methods to estimate and update the model parameters and quantile forecasts through a rolling window approach. Overall, our findings strongly support the validity of our trading strategies and show strong potential as a functional tool that can help practitioners and investors achieve more efficient portfolio allocations.

This study particularly demonstrates the validity of the ST-GARCH model with Student's t-errors and quantile forecasting for the pair trading. This approach works well in practice, because the model itself includes a broad class of financial time series models, capturing very well the characteristics of assets' liquidity, and employs the quantile forecasting. The quantile method provides more robustness in terms of insensitivity to outliers and model bias than the conventional methods and merits itself as a device to control trading strategies feasibly via choosing different quantiles. Our proposed method is a functional tool with sophisticated and versatile applications, that are quite useful in market trading.

While we restricted our interests to stock prices in this study, our method can be applied to other assets and commodities. We leave this issue as our future research project.

Acknowledgements

We thank the Editor, the Associate Editor, and anonymous referees for their valuable time and careful comments, which have improved this paper. Cathy W.S. Chen's research is funded by the Ministry of Science and Technology, Taiwan (MOST 104–2410-H-035–004). Sangyeol Lee's research is supported by the Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (No. 2015R1A2A2A010003894).

Appendix A. 36 stocks

The companies are 3 M (MMM), Alcoa (AA), Apple (AAPL), American Express (AXP), AT&T (T), Bank of America (BAC), Boeing (BA), Caterpillar (CAT), Chevron Corporation (CVX), Cisco Systems (CSCO), Coca-Cola (KO), Dupont (DD), ExxonMobil (XOM), General Electric (GE), Google (GOOGL), The Goldman Sachs Group (GS), Hewlett–Packard (HPQ), The Home Depot (HD), Intel (INTC), IBM (IBM), Johnson & Johnson (JNJ), JPMorgan Chase (JPM), Lowe's Companies (LOW), McDonalds (MCD), Merck (MRK), Microsoft (MSFT), PepsiCo (PEP), Pfizer (PFE), Procter & Gamble (PG), Travelers (TRV), UnitedHealth Group Incorporated (UNH), United Technologies Corporation (UTX), Verizon Communications (VZ), Wal-Mart (WMT), Walt Disney (DIS), and Yahoo (YHOO).

References

Anderson, H. M., Nam, K., & Vahid, F. (1999). Asymmetric nonlinear smooth transition GARCH models. In P. Rothman (Ed.), *Nonlinear Time Series Analysis of Economic and Financial Data* (pp. 191–207). Kluwer.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 31*, 307–327.

Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). ARCH modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics, 52*, 5–59.

Chan, K. S., & Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis, 7*, 179–190.

Chen, C. W. S., Chen, M., & Chen, S. Y. (2014). Pairs trading via three-regime threshold autoregressive GARCH models. In Huynh et al. (Eds.), *Modeling Dependence in Econometrics, Advances in Intelligent Systems and Computing* (pp. 127–140). Switzerland: Springer International Publishing.

Chen, C. W. S., Weng, M. M. C., & Watanabe, T. (2016) Bayesian forecasting of Value-at-Risk based on variant smooth transition heteroskedastic models. Technique report.

Chen, C. W. S., & So, M. K. P. (2006). On a threshold heteroscedastic model. *International Journal of Forecasting, 22*, 73–89.

Elliott, R. J., van der Hoek, J., & Malcolm, W. P. (2005). Pairs trading. *Quantitative Finance, 5*, 271–276.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of variance of United Kingdom inflation. *Econometrica, 50*, 987–1008.

Gatev, E., Goetzmann, W. N., & Rouwenhorst, K. G. (2006). Pairs trading: performance of a relative value trading arbitrage rule. *The Review of Financial Studies, 19*, 797–827.

George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association, 88*, 881–889.

Gerlach, R., & Chen, C. W. S. (2008). Bayesian inference and model comparison for asymmetric smooth transition heteroskedastic models. *Statistics and Computing, 18*, 391–408.

Girma, P. B., & Paulson, A. S. (1999). Risk arbitrage opportunities in petroleum futures spreads. *Journal of Futures Markets, 19*, 931–955.

Granger, C. W. J., & Teräsvirta, T. (1993). *Modelling nonlinear economic relationships*. Oxford: Oxford University Press.

Hastings, W. K. (1970). Monte-Carlo sampling methods using Markov chains and their applications. *Biometrika, 57*, 97–109.

Jansen, E. S., & Teräsvirta, T. (1996). Testing parameter constancy and super exogeneity in econometric equations. *Oxford Bulletin of Economics and Statistics, 58*, 735–768.

Jorion, O. (1997). *Value at risk: The new benchmark for controlling market risk*. Chicago: Irwin Professional.

Krauss, C. (2016). Statistical arbitrage pairs trading strategies: review and outlook. Journal of Economic Surveys. http://dx.doi.org/10.1111/joes.12153.

Liu, J., & Timmermann, A. (2013). Optimal convergence trade strategies. *Review of Financial Studies, 26*, 1048–1086.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics, 21*, 1087–1091.

Perlin, M. S. (2009). Evaluation of pairs-trading strategy at the Brazilian financial market. *Journal of Derivatives and Hedge Funds, 15*, 122–136.

Ross, S. (1976). The arbitrage theory of capital asset pricing. *Journal of Economic Theory, 13*, 341–360.

Teräsvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association, 89*, 208–218.

van Dijk, D., Teräsvirta, T., & Franses, P. H. (2002). Smooth transition autoregressive models – a survey of recent developments. *Econometric Reviews, 21*, 1–47.

Vidyamurthy, G. (2004). *Pairs trading: Quantitative method and analysis*. Hoboken, New Jersey: John Wiley and Sons.